

Exact Inference on Gaussian Graphical Models of Arbitrary Topology using Path-Sums

P.-L. Giscard

P.GISCARD1@PHYSICS.OX.AC.UK

Department of Physics

University of Oxford

Clarendon Laboratory, Oxford OX1 3PU, United Kingdom

Z. Choo

ZHENG.CHOO@STATS.OX.AC.UK

Department of Statistics

University of Oxford

1 South Parks Road, Oxford OX1 3TG, UK

S. J. Thwaite

SIMON.THWAITE@PHYSIK.UNI-MUENCHEN.DE

Department of Physics and Arnold Sommerfeld Center for Theoretical Physics,

Ludwig-Maximilians-Universität München

Theresienstraße 37, 80333 Munich, Germany

D. Jaksch

D.JAKSCH1@PHYSICS.OX.AC.UK

Department of Physics

University of Oxford

Clarendon Laboratory, Oxford OX1 3PU, United Kingdom.

Centre for Quantum Technologies

National University of Singapore

3 Science Drive 2, Singapore 117543

Editor:

Abstract

We present the path-sum formulation for exact statistical inference of marginals on Gaussian graphical models of arbitrary topology. The path-sum formulation gives the covariance between each pair of variables as a branched continued fraction of finite depth and breadth. Our method originates from the closed-form resummation of infinite families of terms of the walk-sum representation of the covariance matrix. We prove that the path-sum formulation always exists for models whose covariance matrix is positive definite: i.e. it is valid for both walk-summable and non-walk-summable graphical models of arbitrary topology. We show that for graphical models on trees the path-sum formulation is equivalent to Gaussian belief propagation. We also recover, as a corollary, an existing result that uses determinants to calculate the covariance matrix. We show that the path-sum formulation is valid for arbitrary partitions of the inverse covariance matrix. We give detailed examples demonstrating our results.

Keywords: Gaussian graphical models, belief propagation, path-sum, walk-sum, graphs of arbitrary topology, block matrices

1. Introduction

A Gaussian Markov random field (GMRF) is a random vector that follows a multivariate normal (or Gaussian) distribution and satisfies *conditional independence* assumptions, hence the *Markov* property. If X_1, X_2, X_3 are random variables with a joint probability density function (or joint probability mass function in a discrete case), we say that X_1 is *conditionally independent* of X_2 given X_3 , denoted $X_1 \perp\!\!\!\perp X_2 | X_3$, if (Lauritzen, 1996)

$$f(x_1, x_2 | x_3) = f(x_1 | x_3) f(x_2 | x_3).$$

Here we use f as a generic symbol for the probability density function of the random variables corresponding to its arguments. GMRFs have a simple interpretation and find their applications, for example, in image analysis, spatial statistics, structural time series analysis and analysis of longitudinal and survival data (Rue and Held, 2005).

Consider a random vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$ following a multivariate normal distribution with mean μ and covariance matrix Σ , denoted $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$. The probability density function of \mathbf{X} is given as

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp \left[-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right].$$

Here Σ is a symmetric and positive definite matrix. We write $A \succ 0$ to denote that the matrix A is positive definite. Alternatively the probability density function of \mathbf{X} can be expressed in a *canonical form*

$$\begin{aligned} f(x) &= g(x) \exp \left[-\frac{1}{2} x^T J x + h^T x - k(\mu, \Sigma) \right] \\ &\propto \exp \left[-\frac{1}{2} x^T J x + h^T x \right], \end{aligned} \tag{1}$$

where $J = \Sigma^{-1}$, $h = J\mu$, $g(x) = (2\pi)^{-\frac{n}{2}}$ and $k(\mu, \Sigma) = -\frac{1}{2} \mu^T J \mu - \frac{1}{2} \ln(\det(J))$. We call J the *information matrix* (or precision matrix) and h the *potential vector*.

One advantage of using the form of parametrization in (1) is that J admits a graphical model in the following sense. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph with the vertex set \mathcal{V} and the set of edges \mathcal{E} . Let $\mathbf{X}_{\setminus ij}$ denote the set of variables with X_i and X_j removed from \mathbf{X} . If $J = (J_{ij})_{i,j \in \mathcal{V}}$ is positive definite, then for $i, j \in \mathcal{V}$, where $i \neq j$, we have (Proposition 5.2, Lauritzen, 1996)

$$X_i \perp\!\!\!\perp X_j | \mathbf{X}_{\setminus ij} \Leftrightarrow J_{ij} = 0.$$

Then we define a GMRF as follows.

Definition 1 (Definition 2.1, Rue and Held, 2005) *A random vector \mathbf{X} is called a GMRF with respect to a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with information matrix J and potential vector h if and only if its density has the form of (1) and $J_{ij} \neq 0 \Leftrightarrow (i, j) \in \mathcal{E}$, for all i and j .*

It is known that X satisfies the Markov property on \mathcal{G} (see Theorem 2.4 by Rue and Held, 2005)¹. Note that by Definition 1, there is a one-to-one correspondence between the structure of J and the structure of \mathcal{G} . Most information matrices J for GMRFs are sparse, i.e. there are $O(n)$ non-zero entries in J . The sparsity structure of J facilitates the simulation of GMRFs through J (Rue and Held, 2005).

Another advantage of using the canonical form concerns the estimation of \mathbf{X} given noisy observations \mathbf{Y} (Johnson, 2002). Indeed, assume that $\mathbf{Y} = C\mathbf{X} + \varepsilon$, where C is an $n \times n$ real matrix and $\varepsilon \sim \mathcal{N}(0, M)$. Then the conditional distribution of $\mathbf{X}|\mathbf{Y}$ is given by

$$\begin{aligned} f(x|y) &= \frac{f(y|x)f(x)}{f(y)} \\ &\propto f(y|x)f(x) \\ &\propto \exp \left[-\frac{1}{2}x^T \tilde{J}x + \tilde{h}^T x \right], \end{aligned}$$

where $\tilde{J} = J + C^T M^{-1}C$ and $\tilde{h} = h + C^T M^{-1}y$. Thus given noisy observations, one only needs to update the information matrix and the potential vector to construct a graphical model for $f(x|y)$. For simplicity, we use J and h to denote the parameters after absorption of observations.

Given the canonical form, one needs both the covariance matrix $\Sigma = J^{-1}$ and the mean vector $\mu = J^{-1}h$ to obtain the marginal distributions of \mathbf{X} or $\mathbf{X}|\mathbf{Y}$. Since knowing J^{-1} is sufficient to recover μ , we focus our efforts on calculating J^{-1} . Direct inversion of J has complexity $O(n^3)$ and does not exploit the sparsity of J .² In a simple situation where the graph \mathcal{G} of a GMRF is a tree, belief propagation (BP) efficiently calculates the correct marginals (Malioutov et al., 2006; Pearl, 1988). For graphs with cycles (also called loopy graphs), the method of loopy belief propagation (LBP) can be used to efficiently *approximate* the marginals. However, it was shown in Weiss and Freeman (2001) that while LBP gives correct means for the marginals, the estimates of the covariance matrices it provides are generally incorrect.

In this article, we present a novel approach to the calculation of the marginals of \mathbf{X} or $\mathbf{X}|\mathbf{Y}$, which we term *method of path-sums*. This approach is a generalization and completion of the walk-sum formulation developed in Malioutov et al. (2006). The method of path-sums is based on results established by Giscard et al. (2012) concerning the algebraic structure of walk sets, which permit the systematic resummation of infinite families of walks in any walk-sum. These resummations transform a walk-sum into a branched continued fraction comprising only a *finite* number of terms. Furthermore, these terms have an elementary interpretation as simple paths and simple cycles on \mathcal{G} . A simple path is a walk (i.e. a trajectory on \mathcal{G}) whose vertices are all distinct. A simple cycle is a walk whose endpoints

-
1. For a GMRF, the pairwise Markov property, the local Markov property and the global Markov property are equivalent. This is proven by using Proposition 3.8 of Lauritzen (1996), in conjunction with the Hammersley-Clifford Theorem (Theorem 3.9, Lauritzen, 1996).
 2. Algorithms that compute *some* entries (either diagonal entries or certain off-diagonal entries) of a symmetric sparse matrix with complexity less than $O(n^3)$ do exist (see Eastwood and Wan, 2013; Li et al., 2008; Lin et al., 2011; Tang and Saad, 2012). The path-sum representation achieves this as well, computing the covariance of a pair of variables with complexity $O(n)$ whenever \mathcal{G} is a tree: see §5 and Giscard et al. (2013).

are identical and intermediate vertices (from the second to the penultimate) are all distinct and different from the endpoints. An important consequence of these observations is that if J is positive definite³, the path-sum formulation of J^{-1} is convergent.

Consequently, one does not need the walk-summability of a model, which was devised by Malioutov et al. (2006) to guarantee the convergence of the infinite walk-sum representation of the covariance matrix $\Sigma = J^{-1}$. Let $R := I - J$ be the partial correlations matrix and $|R|$ be the entrywise absolute value of this matrix, i.e. $(|R|)_{ij} = |R_{ij}|$. The authors showed that a GMRF is walk-summable if and only if the spectral radius $\rho(|R|)$ is strictly less than 1, which implies $J \succ 0$. However, the converse does not hold: that is, there are positive definite information matrices which are not walk-summable. In contrast to the walk-summability criterion, our formulation only requires positive definiteness of J . Most importantly, the path-sum formulation gives the *correct* marginals when the graph \mathcal{G} is loopy, and is equivalent to BP when \mathcal{G} is a tree.

The rest of this article is organized as follows. In §2, we introduce the context and arguments underlying the path-sum formulation. We present the path-sum result in §2.2 and show its validity for all positive definite matrices, irrespective of the walk-summability criterion. In the following section, §3, we relate the path-sum representation of a covariance matrix to existing approaches. In §4 we prove that the path-sum representation can be applied to arbitrary partitions of the information matrix J . We give an example demonstrating this claim. Finally in §5 we briefly discuss the computational cost of our approach as well as future prospects. The proof of the path-sum result is deferred to Appendix A.

2. Path-sum representation

2.1 Context

The “*most basic result of algebraic graph theory*” (as described in Flajolet and Sedgewick, 2009) states that the powers of the adjacency matrix $A_{\mathcal{G}}$ of a graph \mathcal{G} generate all the walks on this graph (Biggs, 1993). This result extends to weighted graphs if $A_{\mathcal{G}}$ is replaced by a weighted adjacency matrix R , with R_{ij} the weight of the edge from vertex j to vertex i on \mathcal{G} . Then $(R^{\ell})_{ij}$ is the sum of the weights of all the walks of length ℓ from j to i (Flajolet and Sedgewick, 2009). The weight of a walk is simply the ordered product of the weight of the edges it traverses. (Note that the indices of a matrix are written right-to-left, but correspond to an edge written left-to-right. This is due to unfortunate conventions). We index the entries of R by the labels of the vertices in \mathcal{G} . We write these labels with roman letters, Greek letters, or numbers, as convenient. Now – assuming that the Taylor series converges – we have $(I - R)^{-1} = \sum_{\ell} R^{\ell}$. It follows that $(I - R)^{-1}_{ij}$ can be interpreted as the sum of the weights of all the walks from j to i . This directly implies the walk-sum interpretation advocated by Malioutov et al. (2006) for $J^{-1} = (I - R)^{-1}$, with $R = I - J$ and \mathcal{G} the graphical model of the GMRF constructed by using Definition 1. Further, it follows that the calculation of J^{-1} is susceptible to a particular resummation technique from graph theory based on the structure of sets of walks, called the method of path-sums.

In its most general form, the method of path-sums stems from a fundamental algebraic property of the set of all walks on any weighted graph: namely, that any walk factorizes

3. $J \succ 0$ is equivalent to J being non-singular for a GMRF.

uniquely into products of prime walks, which are the simple paths and simple cycles of \mathcal{G} (see §1 for the definitions). The path-sum representation of the series of all walks on the graph \mathcal{G} is thus the representation of this series that only involves the prime walks.⁴ Since a finite graph sustains only finitely many primes, the walk series (which is typically infinite) thus has an exact representation involving only finitely many terms. An important consequence of this observation is that the path-sum expression of J^{-1} is *convergent* as long as $J \succ 0$.

For a full exposition of the algebraic structure of walk sets at the origin of path-sums and its applications in linear algebra, the interested reader can refer to Giscard et al. (2012) and Giscard et al. (2013). In §2.2 we give the *explicit* and universal path-sum formulation for J^{-1} . This expression takes the form of a branched continued fraction of finite depth and breadth.

2.2 Path-sum formulation of the covariance matrix

Let $\mathcal{G} \setminus \{\alpha, \beta, \dots\}$ denote the subgraph of \mathcal{G} obtained by deleting from \mathcal{G} the vertices $\{\alpha, \beta, \dots\} \subset \mathcal{V}$ and the edges incident to them. For simplicity, we write $J_{\alpha\beta}^{-1}$ for $(J^{-1})_{\alpha\beta}$. The path-sum expression for $\Sigma = J^{-1}$ is presented in Theorem 2 below. We defer its proof to Appendix A.

Theorem 2 *Let $J \succ 0$ be an information matrix. Let $\Pi_{\mathcal{G}; \alpha\omega}$ and $\Gamma_{\mathcal{G}; \alpha}$ be the sets of simple paths from α to ω on \mathcal{G} and the set of simple cycles from α to itself on \mathcal{G} , respectively. If \mathcal{G} has finitely many vertices and edges, these two sets are finite.*

Then each entry of the covariance matrix $\Sigma = J^{-1}$ admits an expression involving only weighted prime walks, called a path-sum representation. It is explicitly given by

$$J_{\omega\alpha}^{-1} = \sum_{p \in \Pi_{\mathcal{G}; \alpha\omega}} (-1)^{\ell(p)} \prod_{j=1}^{\ell(p)+1} \left\{ (J_{\mathcal{G} \setminus \{\alpha, \nu_2, \dots, \nu_{j-1}\}})^{-1}_{\nu_j \nu_j} J_{\nu_{j+1} \nu_j} \right\} J_{\alpha\alpha}^{-1}, \quad (2)$$

$$J_{\alpha\alpha}^{-1} = \left(\sum_{\gamma \in \Gamma_{\mathcal{G}; \alpha\alpha}} (-1)^{\ell(\gamma)+1} J_{\mu_1 \mu_{\ell(\gamma)}} \prod_{j=2}^{\ell(\gamma)} \left\{ (J_{\mathcal{G} \setminus \{\alpha, \mu_2, \dots, \mu_{j-1}\}})^{-1}_{\mu_j \mu_j} J_{\mu_j \mu_{j-1}} \right\} \right)^{-1}, \quad (3)$$

where the products are right-to-left (i.e. $\prod_{i=1}^m a_i = a_m \cdots a_1$), $p = (\nu_1, \nu_2, \dots, \nu_{\ell(p)+1})$ is a simple path of length $\ell(p)$ with $\alpha \equiv \nu_1$ and $\omega \equiv \nu_{\ell(p)+1}$ for convenience; and $\gamma = (\mu_1, \mu_2, \dots, \mu_{\ell(\gamma)}, \mu_1)$ is a simple cycle of length $\ell(\gamma)$ from $\alpha \equiv \mu_1$ to itself.

Note that $J_{\alpha\alpha}^{-1}$ is obtained recursively through Eq. (3). Indeed it is expressed in terms of entries of inverses of submatrices of J , such as $(J_{\mathcal{G} \setminus \{\alpha, \mu_2, \dots, \mu_{j-1}\}})^{-1}_{\mu_j \mu_j}$, which is in turn obtained through Eq. (3) but on the subgraph $\mathcal{G} \setminus \{\alpha, \dots, \mu_{j-1}\}$ of \mathcal{G} . The recursion stops when vertex μ_j has no neighbor on this subgraph, in which case $(J_{\mathcal{G} \setminus \{\alpha, \mu_2, \dots, \mu_{j-1}\}})^{-1}_{\mu_j \mu_j} = 1/J_{\mu_j \mu_j}$ (note that $J_{\mu_j \mu_j} \neq 0$ since $J \succ 0$). The entry $J_{\alpha\alpha}^{-1}$ is therefore expressed as a branched continued fraction which terminates at a finite depth, and $J_{\alpha\omega}^{-1}$ is a finite sum of such continued fractions.

4. Since path-sums are the prime representations of walk series, they are the graph-theoretic analog of Euler product formulas for the Riemann zeta function and other totally multiplicative functions in number theory.

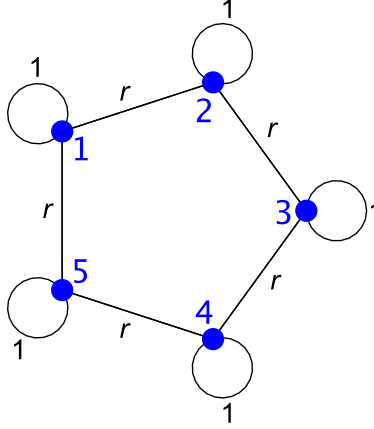


Figure 1: The circle graph on 5 vertices \mathcal{C}_5 associated with J in Example 1. The edge-weights are indicated next to the edges and the vertices are labeled 1 to 5.

Remark 3 Theorem 2 is valid even when \mathcal{G} is loopy and/or J is not walk-summable. In particular there is no restriction on the spectrum of J as long as $J \succ 0$. An example showing this is given below.


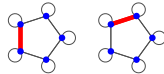
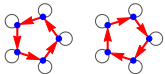
Example 1 (An illustrative example) To illustrate Theorem 2, we consider an example taken from Malioutov et al. (2006), where J has the structure of a circle graph on 5 vertices, denoted \mathcal{C}_5 , see Fig. 1. The information matrix is

$$J = \begin{pmatrix} 1 & r & 0 & 0 & r \\ r & 1 & r & 0 & 0 \\ 0 & r & 1 & r & 0 \\ 0 & 0 & r & 1 & r \\ r & 0 & 0 & r & 1 \end{pmatrix}.$$



Then J is positive definite for $\frac{2}{1-\sqrt{5}} \leq r \leq \frac{2}{1+\sqrt{5}}$ (approximately $-1.618 \leq r \leq 0.618$), and walk-summable if and only if $-1/2 \leq r \leq 1/2$. This example thus provides a test case for both the walk-summable and non-walk-summable situations, depending on the value of r . Here we obtain J^{-1} correctly for all values of r such that J is not singular.

By the symmetry of \mathcal{C}_5 , all the diagonal entries of J^{-1} are identical, and we choose to calculate J_{11}^{-1} without loss of generality. There are five simple cycles from vertex 1 to itself on \mathcal{C}_5 : i) the self-loop $1 \rightarrow 1$; ii) the two backtracks $1 \rightarrow 2 \rightarrow 1$ and $1 \rightarrow 5 \rightarrow 1$; and iii) the two pentagons $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 1$ and $1 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1$. By symmetry, the

two backtracks and the two pentagons have the same weights. Consequently Eq. (3) gives

$$(J^{-1})_{11} = \left(\underbrace{1}_{\text{Self-loop}} - \underbrace{2 r^2 (J_{\mathcal{C}_5 \setminus \{1\}})^{-1}_{22}}_{\text{Backtracks}} + \underbrace{2 r^5 (J_{\mathcal{C}_5 \setminus \{1,2,3,4\}})^{-1}_{55} (J_{\mathcal{C}_5 \setminus \{1,2,3\}})^{-1}_{44} (J_{\mathcal{C}_5 \setminus \{1,2\}})^{-1}_{33} (J_{\mathcal{C}_5 \setminus \{1\}})^{-1}_{22}}_{\text{Pentagons}} \right)^{-1}.$$




The required entries $(J_{\mathcal{C}_5 \setminus \{1\}})^{-1}_{22}, \dots, (J_{\mathcal{C}_5 \setminus \{1,2,3,4\}})^{-1}_{55}$ remain to be calculated. To this end, we use again Eq. (3) of Theorem 2. For example, consider calculating $(J_{\mathcal{C}_5 \setminus \{1\}})^{-1}_{22}$. Since the graph associated with $J_{\mathcal{C}_5 \setminus \{1\}}$ is \mathcal{C}_5 with vertex 1 removed, the only simple cycles from vertex 2 to itself on $\mathcal{C}_5 \setminus \{1\}$ are the self-loop $2 \rightarrow 2$ and the backtrack $2 \rightarrow 3 \rightarrow 2$. We thus find

$$(J_{\mathcal{C}_5 \setminus \{1\}})^{-1}_{22} = \left(\underbrace{1}_{\text{Self-loop}} - \underbrace{r^2 (J_{\mathcal{C}_5 \setminus \{1,2\}})^{-1}_{33}}_{\text{Backtrack}} \right)^{-1},$$



Similarly we obtain $(J_{\mathcal{C}_5 \setminus \{1,2\}})^{-1}_{33} = (1 - r^2 (J_{\mathcal{C}_5 \setminus \{1,2,3\}})^{-1}_{44})^{-1}$, $(J_{\mathcal{C}_5 \setminus \{1,2,3\}})^{-1}_{44} = (1 - r^2 (J_{\mathcal{C}_5 \setminus \{1,2,3,4\}})^{-1}_{55})^{-1}$ and finally $(J_{\mathcal{C}_5 \setminus \{1,2,3,4\}})^{-1}_{55} = 1$. Combining these equations gives

$$\begin{aligned} (J_{\mathcal{C}_5 \setminus \{1,2,3,4\}})^{-1}_{55} = 1 &\Rightarrow (J_{\mathcal{C}_5 \setminus \{1,2,3\}})^{-1}_{44} = \frac{1}{1 - r^2} \\ &\Rightarrow (J_{\mathcal{C}_5 \setminus \{1,2\}})^{-1}_{33} = \frac{1 - r^2}{1 - 2r^2} \\ &\Rightarrow (J_{\mathcal{C}_5 \setminus \{1\}})^{-1}_{22} = \frac{1 - 2r^2}{1 - 3r^2 + r^4}, \\ &\Rightarrow (J^{-1})_{11} = \frac{1 - 3r^2 + r^4}{1 - 5r^2 + 5r^4 + 2r^5}. \end{aligned}$$

As noted, the last equation gives the value of every diagonal entry of J^{-1} . We now consider the off-diagonal entries. We note again that, by the symmetry of \mathcal{C}_5 , there are only two different entries, and we choose to calculate $(J^{-1})_{21}$ and $(J^{-1})_{31}$ without loss of generality. Following Theorem 2, these two entries are given by a sum over simple paths from vertex 1 to vertex 2, and vertex 1 to vertex 3, respectively. In each case there are only two simple paths: for example from 1 to 2, we have $1 \rightarrow 2$ and $1 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2$. Then Eq. (2)

yields

$$\begin{aligned}
 (J^{-1})_{21} = & - \underbrace{r (J_{\mathcal{C}_5 \setminus \{1\}})^{-1}}_{\text{Simple path } \begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \end{array}} (J^{-1})_{11} \\
 & + \underbrace{r^4 (J_{\mathcal{C}_5 \setminus \{1,5,4,3\}})^{-1} (J_{\mathcal{C}_5 \setminus \{1,5,4\}})^{-1} (J_{\mathcal{C}_5 \setminus \{1,5\}})^{-1} (J_{\mathcal{C}_5 \setminus \{1\}})^{-1}}_{\text{Simple path } \begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \end{array}} (J^{-1})_{11} .
 \end{aligned}$$

By the symmetry of \mathcal{C}_5 , we have $(J_{\mathcal{C}_5 \setminus \{1,5,4,3\}})^{-1}_{22} = (J_{\mathcal{C}_5 \setminus \{1,2,3,4\}})^{-1}_{55}$, $(J_{\mathcal{C}_5 \setminus \{1,5,4\}})^{-1}_{33} = (J_{\mathcal{C}_5 \setminus \{1,2,3\}})^{-1}_{44}$, $(J_{\mathcal{C}_5 \setminus \{1,5\}})^{-1}_{44} = (J_{\mathcal{C}_5 \setminus \{1,2\}})^{-1}_{33}$ and $(J_{\mathcal{C}_5 \setminus \{1\}})^{-1}_{55} = (J_{\mathcal{C}_5 \setminus \{1\}})^{-1}_{22}$. The previously obtained results then immediately give

$$(J^{-1})_{21} = \frac{r^4 + 2r^3 - r}{1 - 5r^2 + 5r^4 + 2r^5},$$

and similarly

$$(J^{-1})_{31} = \frac{r^2 - r^3 - r^4}{1 - 5r^2 + 5r^4 + 2r^5}.$$

Piecing these results together, we finally obtain

$$J^{-1} = \frac{r^2}{2r^3 + 3r^2 - r - 1} \begin{pmatrix} \frac{(r-1)r-1}{r^2} & 1 + r^{-1} & -1 & -1 & 1 + r^{-1} \\ 1 + r^{-1} & \frac{(r-1)r-1}{r^2} & 1 + r^{-1} & -1 & -1 \\ -1 & 1 + r^{-1} & \frac{(r-1)r-1}{r^2} & 1 + r^{-1} & -1 \\ -1 & -1 & 1 + r^{-1} & \frac{(r-1)r-1}{r^2} & 1 + r^{-1} \\ 1 + r^{-1} & -1 & -1 & 1 + r^{-1} & \frac{(r-1)r-1}{r^2} \end{pmatrix}.$$

We now verify that this expression is correct in both the walk-summable and non-walk-summable situations:

(a) Walk-summable and $J \succ 0$: set $r = 0.3$. Then to 5 decimal places we have $(J^{-1})_{11} = 1.23975$, $(J^{-1})_{21} = -0.39959$ and $(J^{-1})_{31} = 0.09221$, so that

$$J^{-1} = \begin{pmatrix} 1.23975 & -0.39959 & 0.09221 & 0.09221 & -0.39959 \\ -0.39959 & 1.23975 & -0.39959 & 0.09221 & 0.09221 \\ 0.09221 & -0.39959 & 1.23975 & -0.39959 & 0.09221 \\ 0.09221 & 0.09221 & -0.39959 & 1.23975 & -0.39959 \\ -0.39959 & 0.09221 & 0.09221 & -0.39959 & 1.23975 \end{pmatrix}.$$

One can verify that this result obtained by using the path-sum formulation coincides with the one obtained by direct inversion of J .

(b) **Non walk-summable and $J \succ 0$:** set $r = 0.6$. Then to 5 decimal places, we have $(J^{-1})_{11} = 14.09091$, $(J^{-1})_{21} = -10.90909$ and $(J^{-1})_{31} = 4.09091$, and so

$$J^{-1} = \begin{pmatrix} 14.09091 & -10.90909 & 4.09091 & 4.09091 & -10.90909 \\ -10.90909 & 14.09091 & -10.90909 & 4.09091 & 4.09091 \\ 4.09091 & -10.90909 & 14.09091 & -10.90909 & 4.09091 \\ 4.09091 & 4.09091 & -10.90909 & 14.09091 & -10.90909 \\ -0.39959 & 4.09091 & 4.09091 & -0.39959 & 14.09091 \end{pmatrix}.$$

This result is again easily verified through direct inversion of J .

Remark 4 First, one could have calculated everything with numerical values from the start, as opposed to evaluating the analytic expression of J^{-1} for a specific value of r as we did here. This gives the same results, as expected. Second, the analytical formula for J^{-1} remains valid even when J is not positive definite and only fails for those values of r such that J is singular. The main *mathematical* role of the condition $J \succ 0$ in the path-sum representation is to guarantee that J is not singular.⁵

In the following section, we discuss the relations between Theorem 2 and two existing approaches.

3. Relation to existing approaches

Malioutov et al. (2006) provided a walk-sum derivation for Gaussian belief propagation on trees. Jones and West (2005) presented an expression for the entries of J^{-1} as a sum of simple paths. In this section, we show that these results are corollaries of Theorem 2 arising as special cases.

3.1 Path-sums on trees

The recursive structure of the path-sum representation of J^{-1} is especially simple on trees, for which we recover the Gaussian belief propagation results of Malioutov et al. (2006).

Let $J \succ 0$ be an information matrix associated with a tree model \mathcal{T} . For any vertex α of \mathcal{T} , let $\mathcal{N}(\alpha)$ be the set of neighbors of α on \mathcal{T} . Observe that since \mathcal{T} is a tree, the only simple cycles from α to itself are the self-loop $\alpha \rightarrow \alpha$ with weight $J_{\alpha\alpha}$, and the backtracks to the neighbors of α on \mathcal{T} , e.g. $\alpha \rightarrow \beta \rightarrow \alpha$, $\beta \in \mathcal{N}(\alpha)$. Then Eq. (3) of Theorem 2 gives

$$\Sigma_{\alpha\alpha} = J_{\alpha\alpha}^{-1} = \left(J_{\alpha\alpha} + \sum_{\beta \in \mathcal{N}(\alpha)} -J_{\alpha\beta} (J_{\mathcal{T} \setminus \{\alpha\}})^{-1}_{\beta\beta} J_{\beta\alpha} \right)^{-1}. \quad (4a)$$

The quantity $(J_{\mathcal{T} \setminus \{\alpha\}})^{-1}_{\beta\beta}$ satisfies a similar relation on the subtree $\mathcal{T} \setminus \{\alpha\}$,

$$(J_{\mathcal{T} \setminus \{\alpha\}})^{-1}_{\beta\beta} = \left(J_{\beta\beta} + \sum_{\delta \in \mathcal{N}(\beta) \setminus \alpha} -J_{\beta\delta} (J_{\mathcal{T} \setminus \{\alpha, \beta\}})^{-1}_{\delta\delta} J_{\delta\beta} \right)^{-1}.$$

5. A path-sum result for singular matrices also exists, but it necessitates additional mathematical machinery. In this case the path-sum formulation yields a pseudo-inverse for the singular matrix, see Giscard et al. (2013).

Let $\mathcal{C}_{\mathcal{T} \setminus \{\beta\}; \delta}$ be the connected component of $\mathcal{T} \setminus \{\beta\}$ that contains the vertex $\delta \in \mathcal{N}(\alpha)$. Similarly define $\mathcal{C}_{\mathcal{T} \setminus \{\alpha, \beta\}; \delta}$ to be the connected component of $\mathcal{T} \setminus \{\alpha, \beta\}$ that contains the vertex δ . Since \mathcal{T} is a tree, these components are identical: $\mathcal{C}_{\mathcal{T} \setminus \{\beta\}; \delta} = \mathcal{C}_{\mathcal{T} \setminus \{\alpha, \beta\}; \delta}$, and therefore $(J_{\mathcal{T} \setminus \{\beta\}})_{\delta\delta}^{-1} = (J_{\mathcal{T} \setminus \{\alpha, \beta\}})_{\delta\delta}^{-1}$. Thus, we have

$$(J_{\mathcal{T} \setminus \{\alpha\}})_{\beta\beta}^{-1} = \left(J_{\beta\beta} + \sum_{\delta \in \mathcal{N}(\beta) \setminus \alpha} -J_{\beta\delta} (J_{\mathcal{T} \setminus \{\beta\}})_{\delta\delta}^{-1} J_{\delta\beta} \right)^{-1}. \quad (4b)$$

In order to show that Eqs. (4a, 4b) are the Gaussian belief propagation results, we introduce some notation from Malioutov et al. (2006). Let $\hat{J}_\alpha := 1/\Sigma_{\alpha\alpha}$ and $\hat{J}_{\beta \setminus \alpha}^{-1} := (J_{\mathcal{T} \setminus \{\alpha\}})_{\beta\beta}^{-1}$. With these notations, Eqs. (4a, 4b) become

$$\hat{J}_\alpha = J_{\alpha\alpha} + \sum_{\beta \in \mathcal{N}(\alpha)} \Delta J_{\beta \rightarrow \alpha}, \quad \text{and} \quad \hat{J}_{\beta \setminus \alpha} = J_{\beta\beta} + \sum_{\delta \in \mathcal{N}(\beta) \setminus \alpha} \Delta J_{\delta \rightarrow \beta},$$

with

$$\Delta J_{\delta \rightarrow \beta} = -J_{\beta\delta} \hat{J}_{\delta \setminus \beta}^{-1} J_{\delta\beta}.$$

These are the Gaussian belief propagation equations (Eqs. 7, 8 and 9 in Malioutov et al., 2006), which immediately imply Propositions 16 and 17 in Malioutov et al. (2006) with the further definitions $r_{\beta\alpha} := R_{\beta\alpha} = -J_{\beta\alpha}$, $\gamma_{\beta \setminus \alpha} := (J_{\mathcal{T} \setminus \{\alpha\}})_{\beta\beta}^{-1}$. Note that $r_{\alpha\beta} = r_{\beta\alpha}$ since J is symmetric.

3.2 An approach using determinants

A determinant-based approach to the calculation of the covariance matrix $\Sigma = J^{-1}$ was demonstrated by Jones and West (2005). Here we show that this result follows from Eq. (2) by using the adjugate formula for the matrix inverse. We then point out the fundamental limitation of determinant-based approaches. We overcome this limitation using the path-sum formulation in §4.

We recall the adjugate formula for matrix inverses:

Proposition 5 (Adjugate formula, Strang, 2005) *Let $M \in \mathbb{C}^{n \times n}$ be a $n \times n$ non-singular complex matrix. Then*

$$(M^{-1})_{ij} = (-1)^{i+j} \frac{\det M_j^i}{\det M}, \quad \text{and in particular} \quad (M^{-1})_{ii} = \frac{\det M^i}{\det M},$$

where M_j^i is the matrix M with its i th column and j th row removed and $M^i = M_i^i$.

To obtain this result we start with Eq. (2), which gives here

$$J_{\omega\alpha}^{-1} = \sum_{p \in \Pi_{\mathcal{G}; \alpha\omega}} (-1)^{\ell(p)} (J_{\mathcal{G} \setminus \{\alpha, \nu_2, \dots, \nu_{\ell(p)}\}})_{\omega\omega}^{-1} J_{\omega\nu_{\ell(p)}} \cdots J_{\nu_3\nu_2} (J_{\mathcal{G} \setminus \{\alpha\}})_{\nu_2\nu_2}^{-1} J_{\nu_2\alpha} (J^{-1})_{\alpha\alpha},$$

with $p = (\alpha, \nu_2, \dots, \nu_{\ell(p)}\omega)$ is a simple path of length $\ell(p)$. Since all the entries of J commute, we can reorganize the weights in the above expression to get

$$J_{\omega\alpha}^{-1} = \sum_{p \in \Pi_{\mathcal{G}; \alpha\omega}} (-1)^{\ell(p)} \phi[p] \times (J_{\mathcal{G} \setminus \{\alpha, \nu_2, \dots, \nu_{\ell(p)}\}})_{\omega\omega}^{-1} \times \cdots \times (J_{\mathcal{G} \setminus \{\alpha\}})_{\nu_2\nu_2}^{-1} \times (J^{-1})_{\alpha\alpha},$$

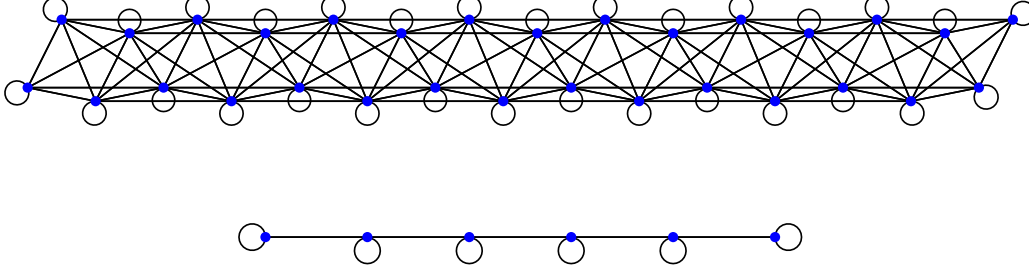


Figure 2: top: the graphical model associated with a 5-banded 30×30 positive definite information matrix. The weights of the edges are the entries of J . Bottom: the graphical model associated with a partition of J into 5×5 blocks. The edge weights are the blocks constituting J .

where $\phi[p] = J_{\omega\nu_{\ell(p)}} \dots J_{\nu_3\nu_2} J_{\nu_2\alpha}$ is the weight of the simple path p . Using the adjugate formula in Proposition 5, we have

$$J_{\omega\alpha}^{-1} = \sum_{p \in \Pi_{\mathcal{G}; \alpha\omega}} (-1)^{\ell(p)} \phi[p] \times \frac{\det J^{\alpha, \nu_2, \dots, \nu_{\ell(p)}, \omega}}{\det J^{\alpha, \nu_2, \dots, \nu_{\ell(p)}}} \times \dots \times \frac{\det J^{\alpha, \nu_2}}{\det J^{\alpha}} \times \frac{\det J^{\alpha}}{\det J},$$

since $J_{\mathcal{G} \setminus \{\alpha, \nu_2, \dots, \nu_j\}} = J^{\alpha, \nu_2, \dots, \nu_j}$. This simplifies to

$$J_{\omega\alpha}^{-1} = \sum_{p \in \Pi_{\mathcal{G}; \alpha\omega}} (-1)^{\ell(p)} \phi[p] \frac{\det J^{\alpha, \nu_2, \dots, \omega}}{\det J}. \quad (5)$$

The last equation is given as Theorem 1 in Jones and West (2005). Note that we only used Eq. (2) in Theorem 2 to obtain this result. Jones and West (2005) did not give an expression equivalent to Eq. (3) in Theorem 2, which provides an explicit formula for the necessary determinants in terms of simple cycles on the graph associated with J .

The requirement that the entries of J be commutative to obtain Eq. (5) from Eq. (2) may seem trivial, but in fact it excludes the very important case of *block matrices*. In the next section, we demonstrate the use of Theorem 2 in the case where J is a block matrix.

4. Path-sums for arbitrary partitions of J

An information matrix J may have a sparsity structure that is best exploited by partitioning J into blocks. For example, consider an $n \times n$ information matrix J which is banded: i.e. $J_{ij} = 0$ if and only if $|i - j| > b$ for some $b < n$. We say that J is b -banded. Then J is simply block-tridiagonal when partitioned $b \times b$ blocks. Consequently, the graph \mathcal{G} associated with this partition of J is simpler than the graph associated with the full matrix: see for example Fig. 2. It is therefore desirable to develop a method capable of exploiting these simplifications.

A fundamental impediment to determinant-based approaches to J^{-1} is that the notion of determinant itself does not extend to matrices with non-commutative blocks (Silvester, 2000). In contrast, the path-sum formulation for J^{-1} does not require the commutativity of the entries of J . For this reason, it continues to hold even when these entries do not commute, see (Giscard et al., 2012) and (Giscard et al., 2013).

Let J be an $n \times n$ information matrix and I_1, \dots, I_B for some $1 \leq B \leq n$ be B disjoint subsets of $\{1, \dots, n\}$. Then we write $J_{I_i I_j}$ for the minor (i.e. block) of J that corresponds to the rows indexed by I_i and the columns indexed by I_j . We form a new graph $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ associated with this partition of J , such that $J_{I_i I_j} \neq \mathbf{0} \iff (j, i) \in \mathcal{E}'$, where $\mathbf{0}$ is a zero matrix. The block $J_{I_i I_j}$ is now the weight associated with the edge (j, i) . With these conventions, Theorem 2 extends to block matrices without modification.

Remark 6 For a GMRF \mathbf{X} , the partition of J given above is equivalent to a partition of the set of random variables \mathbf{X} into $B > 1$ disjoint subsets $\mathbf{X}_1, \dots, \mathbf{X}_B$. Let $\mathbf{X}' = (\mathbf{X}_1, \dots, \mathbf{X}_B)$ be a GMRF with respect to a new graph $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ with information matrix J' . Note that each \mathbf{X}_i is now a random vector instead of a random variable. Following Definition 1:

$$\mathbf{X}_i \perp\!\!\!\perp \mathbf{X}_j | \mathbf{X}'_{\setminus ij} \iff J'_{ij} = \mathbf{0} \iff (j, i) \notin \mathcal{E}', \quad (6)$$

with $\mathbf{X}'_{\setminus ij}$ the set of variables with \mathbf{X}_i and \mathbf{X}_j removed from \mathbf{X}' . Note that $\mathbf{X}_i \perp\!\!\!\perp \mathbf{X}_j | \mathbf{X}'_{\setminus ij}$ implies the global Markov property (GMP), which, for GMRFs, is equivalent to the pairwise Markov property (PMP) (Lauritzen, 1996). However for a general distribution, the PMP does not imply the GMP (Lauritzen, 1996). Hence in this case, the partition of \mathbf{X} as per Eq. (6) may not correspond to the partition of J .

Below we give a detailed synthetic example demonstrating the path-sum formulation of the covariance matrix $\Sigma = J^{-1}$ using a partition of J .

Example 2 (Non walk-summable block matrix on a loopy graph) Consider the following positive definite information matrix J of a thin membrane model

$$J = \begin{pmatrix} a+5b & -b & 0 & -b & -b & 0 & -b & -b & 0 \\ -b & a+5b & -b & -b & 0 & -b & -b & 0 & -b \\ 0 & -b & a+5b & 0 & -b & -b & 0 & -b & -b \\ -b & -b & 0 & a+5b & -b & 0 & -b & -b & 0 \\ -b & 0 & -b & -b & a+5b & -b & -b & 0 & -b \\ 0 & -b & -b & 0 & -b & a+5b & 0 & -b & -b \\ -b & -b & 0 & -b & -b & 0 & a+5b & -b & 0 \\ -b & 0 & -b & -b & 0 & -b & -b & a+5b & -b \\ 0 & -b & -b & 0 & -b & -b & 0 & -b & a+5b \end{pmatrix}, \quad (7)$$

where $a, b > 0$. Recall that walk-summability is equivalent to the condition $\rho(|R|) < 1$, where $|R|$ is the entry-wise absolute value of $R := I - J$ (Malioutov et al., 2006). Here we have $\rho(|R|) = |a + 5b - 1| + \sqrt{19}b + b$, and thus – depending on the values of a and b – walk-summability does not always hold. For example, when $a = b = 1$ we have $\rho(|R|) \simeq 10.36$, $\rho(J) \simeq 9.36$ and $\rho(R) \simeq 8.36$ and J is not walk-summable.

The graph \mathcal{G} associated to J is shown on the left of Fig. 3. Instead of working on this complicated graph, we may partition the matrix into 3×3 blocks as follows

$$J = \begin{pmatrix} L & E & E \\ E & L & E \\ E & E & L \end{pmatrix}, \quad (8)$$

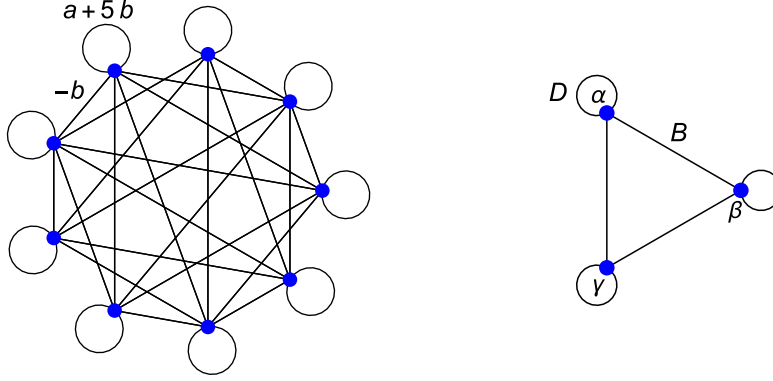


Figure 3: left: graphical model associated with the thin membrane model matrix J of Eq. (7). Every edge has weight $-b$ and every self-loop has weight $a + 5b$. Right: triangular graphical model \mathcal{K}_3 associated with the partition of J given in Eq. (8). Edges and self-loops have matrix weights E and L , respectively. The vertices are labeled α , β and γ for convenience.

with


$$E := \begin{pmatrix} -b & -b & 0 \\ -b & 0 & -b \\ 0 & -b & -b \end{pmatrix}, \quad L := \begin{pmatrix} a + 5b & -b & 0 \\ -b & a + 5b & -b \\ 0 & -b & a + 5b \end{pmatrix}.$$

The graph \mathcal{G}' associated with this partition of J is a triangle, denoted \mathcal{K}_3 and shown on the right of Fig. 3. Accordingly, the edge-weights are now the matrices E and L for edges and self-loops, respectively. Note that these weights do not commute: $[E, L] = EL - LE \neq 0$.

The only simple cycles from any vertex (say, α) to itself on this graph are the self-loop $\alpha \rightarrow \alpha$, the two backtracks $\alpha \rightarrow \beta \rightarrow \alpha$ and $\alpha \rightarrow \gamma \rightarrow \alpha$ and the two triangles $\alpha \rightarrow \beta \rightarrow \gamma \rightarrow \alpha$ and $\alpha \rightarrow \gamma \rightarrow \beta \rightarrow \alpha$. Theorem 2 therefore yields

$$J_{\alpha\alpha}^{-1} = \left(\underbrace{L}_{\text{Self loop}} - \underbrace{E \cdot (J_{\mathcal{K}_3 \setminus \{\alpha\}})^{-1}_{\beta\beta} \cdot E}_{\text{Backtrack}} - \underbrace{E \cdot (J_{\mathcal{K}_3 \setminus \{\alpha\}})^{-1}_{\gamma\gamma} \cdot E}_{\text{Backtrack}} + \underbrace{E \cdot (J_{\mathcal{K}_3 \setminus \{\alpha, \beta\}})^{-1}_{\gamma\gamma} \cdot E \cdot (J_{\mathcal{K}_3 \setminus \{\alpha\}})^{-1}_{\beta\beta} \cdot E}_{\text{Triangle}} + \underbrace{E \cdot (J_{\mathcal{K}_3 \setminus \{\alpha, \gamma\}})^{-1}_{\beta\beta} \cdot E \cdot (J_{\mathcal{K}_3 \setminus \{\alpha\}})^{-1}_{\gamma\gamma} \cdot E}_{\text{Triangle}} \right)^{-1}, \quad (9a)$$


where, by Eq. (3),

$$(J_{\mathcal{K}_3 \setminus \{\alpha\}})^{-1}_{\beta\beta} = \left(\underbrace{L}_{\text{Self loop}} - \underbrace{E \cdot (J_{\mathcal{K}_3 \setminus \{\alpha, \beta\}})^{-1}_{\gamma\gamma} \cdot E}_{\text{Backtrack}} \right)^{-1} \text{ and } (J_{\mathcal{K}_3 \setminus \{\alpha, \beta\}})^{-1}_{\gamma\gamma} = \underbrace{L^{-1}}_{\text{Self loop}}.$$

(9b)

Since $(J_{\mathcal{K}_3 \setminus \{\alpha\}})^{-1}_{\gamma\gamma} = (J_{\mathcal{K}_3 \setminus \{\alpha\}})^{-1}_{\beta\beta}$ and $(J_{\mathcal{K}_3 \setminus \{\alpha, \gamma\}})^{-1}_{\beta\beta} = (J_{\mathcal{K}_3 \setminus \{\alpha, \beta\}})^{-1}_{\gamma\gamma}$, Eqs. (9a, 9b) completely determine $J_{\alpha\alpha}^{-1}$. Thus, we have

$$J_{\alpha\alpha}^{-1} = \frac{b}{a^2 + 8ab - 3b^2} \times \begin{pmatrix} \frac{a}{b} + \frac{a}{a+3b} + \frac{5a}{3(a+6b)} + \frac{1}{3} & 1 & \frac{3b}{a+3b} \\ 1 & \frac{a}{b} + \frac{12a}{5(a+5b)} + \frac{3}{5} & 1 \\ \frac{3b}{a+3b} & 1 & \frac{a}{b} + \frac{a}{a+3b} + \frac{5a}{3(a+6b)} + \frac{1}{3} \end{pmatrix}.$$

By symmetry of J , this gives also $J_{\beta\beta}^{-1}$ and $J_{\gamma\gamma}^{-1}$. The off-diagonal blocks, which are all identical because of the symmetry of the matrix, now follow easily. For example, we have

$$(J^{-1})_{\beta\alpha} = \underbrace{(J_{\mathcal{K}_3 \setminus \{\alpha\}})^{-1}_{\beta\beta} \cdot E \cdot (J^{-1})_{\alpha\alpha}}_{\text{Path}} - \underbrace{E \cdot (J_{\mathcal{K}_3 \setminus \{\alpha, \gamma\}})^{-1}_{\beta\beta} \cdot E \cdot (J_{\mathcal{K}_3 \setminus \{\alpha\}})^{-1}_{\gamma\gamma} \cdot E \cdot (J^{-1})_{\alpha\alpha}}_{\text{Path}},$$


$$= \frac{b}{a^2 + 8ab - 3b^2} \begin{pmatrix} \frac{a}{a+3b} - \frac{5a}{6(a+6b)} + \frac{5}{6} & 1 & \frac{3b}{a+3b} \\ 1 & \frac{6b}{a+5b} & 1 \\ \frac{3b}{a+3b} & 1 & \frac{a}{a+3b} - \frac{5a}{6(a+6b)} + \frac{5}{6} \end{pmatrix}.$$

As long as $a, b > 0$ (which implies $J \succ 0$), these results hold regardless of the values of a and b . For example setting $a = b = 1$, we obtain the correct covariance matrix even though J is not walk-summable.

5. Conclusion

We have presented an approach for exact inference of marginals in Gaussian graphical models, called the path-sum formulation. This approach represents the covariance matrix Σ of a Gaussian Markov Random Field (GMRF) as a continued fraction of finite depth and breadth, and which only involves the simple paths and simple cycles on the graphical model of the GMRF. For GMRFs, we have shown that the path-sum formulation only requires $J \succ 0$. It also extends to arbitrary partitions of J , thereby providing the flexibility necessary to exploit the sparsity structure of the information matrix $J = \Sigma^{-1}$.

From an algorithmic point of view, the computational cost associated with the calculation of any entry of a $n \times n$ covariance matrix using a path-sum is known to be $O(n)$ on trees,

see Giscard et al. (2013). The $O(n)$ complexity is achieved for the most general situation; if J or \mathcal{G} has symmetries, the complexity would be lower. Evaluating the complexity of the path-sum approach on a graph with an arbitrary topology remains an open problem. Since the method of path-sums arises from exact resummations on walk-sums, every problem that has a walk-sum interpretation is susceptible to these resummations and therefore admits a path-sum expression. Consequently, every algorithm that has a walk-sum interpretation (see e.g. Chandrasekaran et al., 2008) necessarily has a path-sum formulation.

The interpretation of the entries of the covariance matrix as walk-sums and the existence of the path-sum formulation opens the door to many more walk-based methods, which are intermediary between these two approaches. The central idea is to exploit recent results regarding the algebraic structure of the set of walks on graphs with arbitrary topology (Thwaite, 2014) to identify certain infinite geometric series of terms appearing in a walk-sum. These geometric series can then be exactly resummed, thereby reducing the sum of all walk weights to a sum over the weights of a certain (yet infinite) subset of ‘irreducible’ walks. Each term in this sum is ‘dressed’ so as to exactly include the contributions of the infinite families of resummed terms. The exact form of both the dressing and the irreducible terms remaining in the sum depend on the structure of the resummed terms: choosing a different family of terms produces a different series.

Viewed in this context, the walk-sum and path-sum formulations of a given problem can be seen to be the extrema of a hierarchy of possible resummations: a walk-sum corresponds to the case where *no* terms are exactly resummed, such that an explicit summation over all walks remains to be carried out, while the path-sum expression corresponds to the case where *all possible* geometric series have been resummed, leaving behind only a (finite) sum over simple paths. In between these extremes lie many intermediate formulations, whose mathematical properties such as complexity and convergence are expected to interpolate between those of walk-sums and path-sums.

Acknowledgments

P.-L. Giscard and D. Jaksch acknowledge funding from EPSRC Grant EP/K038311/1. D. Jaksch also received funding from the ERC under the European Unions Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement no. 319286 Q-MAC. Z. Choo was funded by a departmental studentship of the Department of Statistics, University of Oxford. S. J. Thwaite acknowledges funding from the Alexander von Humboldt Foundation.

Appendix A.

In this appendix we prove Theorem 2.

Proof The proof of the theorem is organized in two steps. First, assuming walk-summability, we obtain the path-sum formulation for J^{-1} . Second, we show that path-sum expression thus obtained is the unique analytic continuation of the sum of all walks, and continues to exist in the absence of walk-summability. Consequently, the path-sum remains a valid representation of J^{-1} when walk-summability fails.

Step 1: the path-sum expression for J^{-1} is a special case of the more general result concerning the path-sum formulation of the matrix inverse function presented and proved in Giscard et al. (2013). Below we briefly recount how we obtain result for the specific case of J^{-1} . Let $R := I - J$ and $|R|$ be the entry-wise absolute value of R , i.e. $|R|_{ij} = |R_{ij}|$. Let $\rho(|R|)$ be the spectral radius of $|R|$, i.e. the largest eigenvalue of $|R|$. As established by Malioutov et al. (2006), $\rho(|R|) < 1$ is equivalent to walk-summability, and is enough to guarantee absolute convergence of the series $\sum_{n \geq 0} R^n = (I - R)^{-1} = J^{-1}$. This power series can be seen as a sum of walk weights on the graph \mathcal{G} associated with R (Flajolet and Sedgewick, 2009), that is

$$\left(\sum_{n \geq 0} R^n \right)_{\omega\alpha} = \sum_{w \in \mathcal{W}_{\mathcal{G}; \alpha\omega}} \phi[w], \quad (10)$$

with $\phi[w]$ the weight of the walk w , which is defined to be the product of the weights of the edges traversed by w , where the weight of an edge from α to β is given by

$$\phi[\alpha\beta] := R_{\beta\alpha} = \begin{cases} -J_{\beta\alpha} & \text{if } \alpha \neq \beta, \\ 1 - J_{\alpha\alpha} & \text{if } \alpha = \beta. \end{cases} \quad (11)$$

Note that since we assumed walk-summability, the right-hand side of Eq. (10) exists. We then use the result by Giscard et al. (2012), which reduces a series of weighted walks, such as the one of Eq. (10), to a sum of weighted simple paths and simple cycles. This result is reproduced here for the sake of completeness:

Theorem 7 (Path-sum, Giscard et al., 2012) *Let \mathcal{G} be a graph and $\phi[\cdot]$ be the weight function of Eq. (11). Suppose that the walk-sum $\sum_{w \in \mathcal{W}_{\mathcal{G}; \alpha\omega}} \phi[w]$ exists. Then this sum is given by the weighted path-sum*

$$\sum_{w \in \mathcal{W}_{\mathcal{G}; \alpha\omega}} \phi[w] = \sum_{p \in \Pi_{\mathcal{G}; \alpha\omega}} \prod_{j=1}^{\ell(p)+1} \left\{ \phi_{\mathcal{G} \setminus \{\alpha, \nu_2, \dots, \nu_{j-1}\}; \nu_j} \phi[\nu_{j+1} \nu_j] \right\} \phi_{\mathcal{G}; \alpha},$$

where $\phi_{\mathcal{G}; \alpha} := \sum_{w \in \mathcal{W}_{\mathcal{G}; \alpha\alpha}} \phi[w]$ is the weighted sum of all walks from α to itself on \mathcal{G} and is explicitly given by

$$\phi_{\mathcal{G}; \alpha} = \left(1 - \sum_{\gamma \in \Gamma_{\mathcal{G}; \alpha\alpha}} \phi[\mu_1 \mu_{\ell(\gamma)}] \prod_{j=2}^{\ell(\gamma)} \left\{ \phi_{\mathcal{G} \setminus \{\alpha, \mu_2, \dots, \mu_{j-1}\}; \mu_j} \phi[\mu_j \mu_{j-1}] \right\} \right)^{-1},$$

and similarly for all $\phi_{\mathcal{G} \setminus \{\alpha, \nu_2, \dots, \nu_{j-1}\}; \nu_j}$ and $\phi_{\mathcal{G} \setminus \{\alpha, \mu_2, \dots, \mu_{j-1}\}; \mu_j}$. In these expressions, the products are right-to-left, $p = (\nu_1, \nu_2, \dots, \nu_{\ell(p)+1})$ is a simple path of length $\ell(p)$ with $\alpha \equiv \nu_1$ and $\omega \equiv \nu_{\ell(p)+1}$ for convenience; and $\gamma = (\mu_1, \mu_2, \dots, \mu_{\ell(\gamma)}, \mu_1)$ is a simple cycle of length $\ell(\gamma)$ from $\alpha \equiv \mu_1$ to itself.

By using the edge weights of Eq. (11), we obtain $\phi_{\mathcal{G}; \alpha} = J_{\alpha\alpha}^{-1}$ and similarly $\phi_{\mathcal{G} \setminus \{\alpha, \nu_2, \dots, \nu_{j-1}\}; \nu_j} = (J_{\mathcal{G} \setminus \{\alpha, \nu_2, \dots, \nu_{j-1}\}})_{\nu_j \nu_j}^{-1}$ and Theorem 7 yields Eqs. (2, 3). Next we prove that Eqs. (2, 3) yield J^{-1} for any matrix $J \succ 0$, even when walk-summability does not hold and the walk-sum of Eq. (10) does not converge. Central to our proof is the (well-established) theory of analytic continuation (Priestley, 2003).

Step 2: we consider the three following functions of a complex variable z into $\mathbb{C}^{n \times n}$, $g_1(z) := (I - zR)^{-1}$, $g_2(z) := \sum_{n \geq 0} z^n R^n$ and, since $g_2(z)$ is a walk-sum, it has a path-sum expression which we denote $g_3(z)$. (The path-sum expression $g_3(z)$ is obtained from Theorem 2 on using the edge-weights $zR_{\beta\alpha}$ for an edge from α to β).

The first function, $g_1(z)$, is analytic on $z \in \mathbb{C} \setminus \text{Sp}^{-1}(R)$ with $\text{Sp}^{-1}(R)$ the inverse of the spectrum of R . The second function, $g_2(z)$, is analytic on the disk D_R of the complex plane where $z < 1/\rho(|R|)$. The third function, $g_3(z)$, comprises only a finite number of terms (since there are finitely many simple paths and simple cycles on a finite graph). Consequently $g_3(z)$ exists for $z \in \mathbb{C} \setminus \text{Sp}^{-1}(R)$. Evidently g_1 and g_2 agree on D_R and by construction g_2 and g_3 agree on D_R as well. It follows that g_1 and g_3 constitute two *direct analytic continuations* of g_2 outside of D_R (Priestley, 2003) (also called *extensions* of g_2 , Kreyszig, 1989). By the Uniqueness Theorem (Theorem 15.9, Priestley, 2003) only one such analytic continuation exists and $g_1(z) = g_3(z)$ on the domain $\mathbb{C} \setminus \text{Sp}^{-1}(R)$.

We conclude the proof by showing that the point $z = 1$, for which $g_1(z)$ and $g_3(z)$ yield J^{-1} , is also in this domain. Indeed, the covariance matrix Σ of a Gaussian distribution must be positive definite, implying that it is non-singular (Corollary 7.17 in Horn and Johnson, 2013). Consequently, J is not singular: so J^{-1} exists and 1 is not an eigenvalue of $R = I - J$, i.e. $1 \in \mathbb{C} \setminus \text{Sp}^{-1}(R)$. Then g_1 and g_3 exist at $z = 1$, in particular $g_3(1)$ is a valid representation of J^{-1} , even though $z = 1$ may not be in D_R (i.e. J is not walk-summable). This completes the proof of Theorem 2. ■

References

- N. Biggs. *Algebraic Graph Theory*. Cambridge University Press, Cambridge, 2nd edition, 1993.
- V. Chandrasekaran, J. K. Johnson, and A. S. Willsky. Estimation in Gaussian graphical models using tractable subgraphs: a walk-sum analysis. *IEEE Transactions on Signal Processing*, 56:1916–1930, 2008.
- S. Eastwood and J. W. L. Wan. Finding off-diagonal entries of the inverse of a large symmetric sparse matrix. *Numerical Linear Algebra with Applications*, 20:74–92, 2013.
- P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, Cambridge, 1st edition, 2009.
- P.-L. Giscard, S. J. Thwaite, and D. Jaksch. Continued fractions and unique factorization on digraphs. *arXiv:1202.5523*, 2012.
- P.-L. Giscard, S. J. Thwaite, and D. Jaksch. Evaluating matrix functions by resummations on graphs: the method of path-sums. *SIAM Journal on Matrix Analysis & Applications*, 34:445–469, 2013.
- R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2nd edition, 2013.

- J. Johnson. Walk-summable Gauss-Markov random fields, 2002. unpublished manuscript, available at <http://ssg.mit.edu/~jasonj/johnson-walksum-tr02.pdf>.
- B. Jones and M. West. Covariance decomposition in undirected Gaussian graphical models. *Biometrika*, 92:779–786, 2005.
- E. Kreyszig. *Introductory Functional Analysis with Applications*. John Wiley & Sons, New York; London, revised edition, 1989.
- S. L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.
- S. Li, S. Ahmed, G. Klimeck, and E. Darve. Computing entries of the inverse of a sparse matrix using the FIND algorithm. *Journal of Computational Physics*, 227:9408–9427, 2008.
- L. Lin, C. Yang, J. C. Meza, L. Ying J. Lu, and W. E. SelInv - An algorithm for selected inversion of a sparse matrix. *ACM Transactions on Mathematical Software*, 37:1–19, 2011.
- D. M. Malioutov, J. K. Johnson, and A. S. Willsky. Walk-sums and belief propagation in Gaussian graphical models. *Journal of Machine Learning Research*, 7:2031–2064, 2006.
- J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Mateo, California, 1988.
- H. A. Priestley. *Introduction to Complex Analysis*. Oxford University Press, 2nd edition, 2003.
- H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall/CRC, 2005.
- J. R. Sylvester. Determinants of block matrices. *The Mathematical Gazette*, 84:460–467, 2000.
- G. Strang. *Linear Algebra and Its Applications*. Brooks Cole, 4th edition, 2005.
- J. M. Tang and Y. Saad. A probing method for computing the diagonal of a matrix inverse. *Numerical Linear Algebra with Applications*, 19:485–501, 2012.
- S. J. Thwaite. A family of partitions of the set of walks on a directed graph. *arXiv:1409.3555*, 2014.
- Y. Weiss and W. T. Freeman. Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Neural Computation*, 13:2173–2200, 2001.